Applied Cognitive Psychology

WILEY

SPECIAL ISSUE ARTICLE OPEN ACCESS

The Effects of Acute Alcohol Intoxication on Metamemory Processes and Accuracy When Recalling a Rape Scenario

Madeleine P. Ingham[1] | Brittany D. Gibbs[1] | Melissa F. Colloff[1] | Laura M. Stevens[1,2] | Orli M. Edwards[1] | Sarah R. Rockowitz[1] | Rumandeep K. Hayre[1,3] | Mussaffa Butt[4] | Chloe A. Morris[1] | Heather D. Flowe[1]

[1]School of Psychology, University of Birmingham, Birmingham, United Kingdom | [2]Institute for Social Justice & Crime, University of Suffolk, Ipswich, United Kingdom | [3]Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom | [4]Department of Psychology, Government College University, Lahore, Pakistan

Correspondence: Heather D. Flowe (h.flowe@bham.ac.uk)

Received: 7 September 2024 | **Revised:** 31 January 2025 | **Accepted:** 14 February 2025

Funding: This work was supported by the Economic and Social Research Council (ES/J005169/1), and a British Academy Mid-Career Fellowship Award to H.D.F.

Keywords: alcohol | eyewitness memory | metamemory | police interview | recall accuracy | sexual assault

ABSTRACT

This study examines how acute alcohol intoxication during the encoding of a rape scenario affects metamemory processes and recall accuracy during police interviews. Metamemory is the ability to monitor and control memory reporting. We conducted a secondary data analysis of interview transcripts, applying a novel analytical approach to capture metamemory processes. Twenty-two women aged 18–25 ($M = 20.00$, $SD = 1.63$) were randomly assigned to be either intoxicated or sober during scenario encoding but sober during recall when they underwent a cognitive interview 1 week later. Accuracy was significantly lower in the question compared to free recall phase, particularly in the alcohol condition. Inaccurate recall was preceded by a higher incidence of metamemory indicators of increased retrieval effort (pauses, hedges, fillers), particularly in the question phase for intoxicated participants. These findings elucidate the effects of alcohol on metamemory and memory reporting during police interviews and highlight the need for research about techniques the police can use to maintain recall accuracy over the entire interview process.

1 | Introduction

Sexual assault victims who were alcohol-intoxicated at the time of the offense face considerable challenges when providing their accounts of what happened in the criminal justice system. Chanel Miller, a survivor of a high-profile sexual assault during which she was alcohol intoxicated, poignantly articulates the challenges faced by victims:

> "My memory loss would be used against me. My testimony was weak, was incomplete, and I was made to believe that perhaps I am not enough to win this. His attorney constantly reminded the jury, the only one we can believe is Brock, because she doesn't remember." (Miller 2019, 348).

Miller's experience highlights a critical issue in the prosecution of sexual offenses. Namely, when victims were alcohol intoxicated during the assault, and have incomplete memories of it, these circumstances are often used by the defense to undermine their credibility (Flowe and Carline 2021). Westera and Powell (2017) further explained that jurors often expect victims to provide highly detailed and consistent accounts. When victims cannot do so, perhaps owing to trauma and alcohol intoxication during the offense, their credibility may be questioned. Westera et al. (2017) noted that even minor inconsistencies in a victim's statements and testimony can be used by defense attorneys to cast doubt on the entirety of the victim's account.

The issue of alcohol and memory is particularly relevant in sexual assault because victims are often intoxicated during the offense (Abbey et al. 2004; Mohler-Kuo et al. 2004). Flowe et al. (2019) experimentally examined how alcohol intoxication during the encoding of a rape scenario affects memory reporting during police interviews. They found that while alcohol-intoxicated participants reported fewer correct details, the accuracy of the reported details was comparable to that of sober participants when freely recalling their memories. However, when answering follow-up questions, intoxicated participants showed a tendency to report more inaccurate information, particularly in response to specific questions about perpetrator appearance and actions, though these differences did not reach statistical significance. Apart from this study, there is no other research specifically about supporting memory retrieval during police interviews with sexual assault victims who were intoxicated during the offense.

Best practice guidelines in eliciting accounts emphasize the importance of using a funnel approach, beginning the interview by using open-ended questions before progressing to more specific ones (e.g., Ministry of Justice 2022; Fisher 2010; College of Policing 2022). In keeping with this, police in the UK and elsewhere use the cognitive interview, a procedure that includes techniques for supporting memory retrieval, such as rapport building and instructions to use cognitive mnemonics (e.g., mental imagery), and uses a phased interview approach, including a free recall phase followed by a questioning phase, wherein the interviewer follows up on specific information. UK sexual assault victims' experiences with this type of police interview are mixed (Stern 2010). Some report positive experiences, feeling they had the opportunity to give a full account, while others found the process lengthy and emotionally draining. Other research has found that sexual assault victims may face skepticism and be challenged about the plausibility of their accounts during the police interview (Antaki et al. 2015; MacLeod 2020). Given the potential impact of interviewing practices on the accuracy and completeness of victim accounts, there is a critical need to study how memory reporting unfolds across different phases of an interview, particularly when alcohol was involved. In the current study, we begin to address this by examining metamemory processes when people freely report their memories versus when they are asked follow-up questions.

Metamemory is a cognitive process encompassing both monitoring and control functions, which plays a crucial role in regulating the accuracy and completeness of memory reports (Koriat and Goldsmith 1996; Nelson and Narens 1994). Individuals engage in monitoring processes to evaluate the quality of remembered information and employ control processes to decide whether to report this information based on situational demands and personal goals (Goldsmith et al. 2002). The information remembered during the monitoring process is reported if its quality exceeds the control threshold. The threshold level is set depending on whether the accuracy of the information reported is more important than its informativeness (i.e., whether the investigation requires a high degree of completeness or granularity of detail) or vice versa.

The interview approach by the police affects metamemory processes. Under free recall conditions, where witnesses respond to open-ended interview prompts without interruption, reported information tends to be highly accurate compared to closed questioning (Fisher 2010). As another example, the cognitive interview has been widely researched. It has been shown to increase the amount of correct details that participant witnesses recall about events, with a small increase in the number of incorrect details reported compared to control interviews (see Memon et al. 2010 for a meta-analysis). Conversely, certain interviewing practices lead participant witnesses to report less accurate information (see Roberts and Higham 2002). Authoritative pressure from interviewers, for example, may cause witnesses to lower their control threshold. As a result, weaker and less accurate memories are volunteered (Goldsmith et al. 2005; Koriat and Goldsmith 1996).

Bringing together research on metamemory and phased interviewing, we propose the theoretical framework illustrated in Figure 1. Our framework posits that interviewees report their strongest memories (i.e., memories that are retrieved quickly and that are rich in detail, Mickes et al. 2013) during free recall. The interviewee will typically maintain a relatively stringent report threshold during free recall and volunteer information only if they are confident the information is accurate.

However, when interviewees are prompted to provide further details during the question phase, they face a dilemma: their pool of unreported memories (see right panel of Figure 1) consists of weaker traces than those in the free recall phase (see left panel of Figure 1); yet, they may feel pressure to report less well-remembered information to be helpful or to avoid appearing uncooperative (see Ackerman and Goldsmith 2008). We propose that this leads to a downward shift in the report threshold, and thus, lower recall accuracy during the question phase compared to the free recall phase, especially for witnesses with weaker memories to begin with, such as individuals who were intoxicated at the time of the offense.

We can further draw on this framework to consider how memory reporting differs for interviewees who were sober versus intoxicated during the crime. While alcohol-intoxicated witnesses tend to provide less complete accounts, their accuracy often matches that of sober counterparts under free report conditions (Jores et al. 2019). However, repeated questioning (Oorsouw and Merckelbach 2012) and repeated suggestive questioning (Hildebrand Karlén et al. 2014) can increase incorrect information reporting for participant witnesses who were intoxicated during the crime.

To examine memory reporting, we sought to capture independent measures of metamemory processes. Therefore, we measured behavioral indicators of cognitive effort, such as pauses and utterances of hedges and filler words, during the interview. Previous experimental work has found such indicators are predictive of recall accuracy. In two experiments ($N = 34$ and $N = 20$ participants), Lindholm et al. (2018) showed participants a mock-crime video, then interviewed them afterwards. Interviews were coded for correct (i.e., statements that contained all correct details) and incorrect statements (i.e., statements that contained all incorrect details) using a coding template containing all details present in the video. The authors categorized indicators into four groups: 1. Hedges; 2. Fillers; 3. Non-word fillers; and 4. Delays/pauses (see Table 1 for operationalisation and examples). They found that the occurrence of hedges, non-word fillers, and delays when recounting details was associated with lower recall accuracy. Building on this work, Gustafsson et al. (2019) investigated metamemory in a sample of 22 participants who were interviewed after they encoded a mock-crime video. They found that statements accompanied by hedges, word fillers, and delays tended to be less accurate. More recently, Gustafsson et al. (2022) examined how retrieval effort indicators relate to accuracy over time and with repeated questioning in 56 participants after they encoded a mock-crime video, finding again that hedges and delays were associated with lower accuracy within and between interviews.

Our analysis extends previous research in several important ways. First, we investigate how metamemory processes vary across different phases of a cognitive interview, providing insight into how retrieval effort may change as witnesses progress from free recall to focused questioning. Similar to other studies of the cognitive interview, Flowe et al. (2019) analyzed only the number of details recalled and their accuracy and did not analyze indicators of metamemory processes. Importantly, our novel secondary data analysis of Flowe et al. (2019) aligns with conversation analytic approaches (e.g., Stokoe 2013), which emphasize the importance of studying *how* information is conveyed in real-world interactions, particularly in institutional settings like police interviews. Second, we investigate behavioral indicators of metamemory processes in the context of recalling a hypothetical sexual assault scenario, an area unexplored in previous metamemory studies. Finally, our inclusion of alcohol as a variable is critical, as it addresses widespread beliefs among experts, laypersons, and law enforcement that intoxication during encoding compromises recall accuracy (e.g., see Davis and Loftus 2015; Evans and Schreiber Compo 2010; Evans et al. 2009).

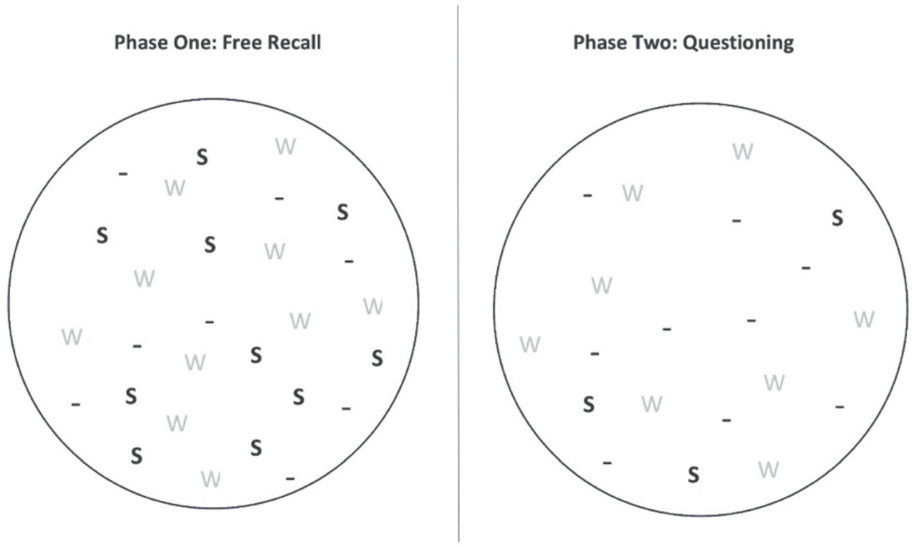

FIGURE 1 | Hypothetical memory search sets during the police interview for a participant who experienced the rape scenario. Note. The search set on the left is for the Free Recall phase and the one on the right is for the Question Phase of the interview. The dashed symbols represent scenario details the participant has forgotten, the S's represent scenario details the participant strongly remembers, and the W's represent scenario details the participant weakly remembers. There are fewer items in the search set for the Question compared to the Free Recall phase because the details the participant reported during the Free Recall phase have been eliminated. Details that are strongly compared to weakly remembered are associated with substantial source information, are recalled more often in the Free Recall compared to the Question Phase, are more likely to be accurate and remembered with high confidence, and are more likely to pertain to central compared to peripheral aspects of the scenario.

TABLE 1 | Operationalisation of the metamemory indicators in the participants' responses (from Lindholm et al. 2018; Gustafsson et al. 2019, 2022).

Measure	Description	Examples
Pauses	Pause of 2 s or longer before or within a response.	—
Non-word fillers	Non-word interjections or sounds that are made during recall effort.	“um”, “hmm”, “erm”, “aaah”
Word fillers	Words or phrases that do not add meaning to the participants' response when used during recall effort. This also includes self-talk.	“you know”, “well”, “like”, “oh”, “let's see”, “oh that wouldn't make sense”
Word hedge	Words or phrases that suggested low confidence or a lack of willing to commit, as well as diminishing the value of an assertion.	“maybe”, “kind of”, “sort of”, “possibly”, “just”, “I think”, “I'm not sure”

2 | Hypotheses

Our study addresses several key hypotheses. First, we hypothesized that recall accuracy is lower in the question phase compared to the free recall phase (Hypothesis 1). This hypothesis is grounded in research showing that witnesses initially report their strongest memories first (e.g., Mickes et al. 2013) leaving predominantly weaker memory traces for the subsequent questioning phase. Second, we hypothesized that participants who were alcohol-intoxicated during encoding exhibit lower recall accuracy during the question phase compared to sober participants (Hypothesis 2). This prediction stems from research that found alcohol-intoxicated compared to sober individuals encode fewer contextual details, resulting in their having a higher proportion of weak memory traces (Söderlund et al. 2005). Third, drawing on recent metamemory research (Gustafsson et al. 2022), we predicted that retrieval effort is negatively associated with recall accuracy (Hypothesis 3). This hypothesis reflects the idea that stronger memories are retrieved more fluently (Mickes et al. 2013). Finally, we hypothesized that the relationship between retrieval effort indicators and recall accuracy is stronger in the question phase compared to the free recall phase, especially for individuals who were intoxicated (Hypothesis 4). This prediction is based on the theoretical framework presented in Figure 1, which posits that interviewer prompts posed during the question phase led the interviewee to lower their threshold and report weaker memories. Since alcohol intoxication at encoding leads to weaker memory formation, information volunteered during the question phase will be less accurate for those who have consumed alcohol compared to those who did not.

3 | Materials and Methods

3.1 | Design and Procedure

A secondary data analysis of Flowe et al. (2019) was undertaken. The unit of analysis was detail, with the dataset containing 923 details (correct $n = 777$, incorrect $n = 146$) from 22 women (tonic $n = 10$, alcohol intoxicated $n = 12$) aged between 18 and 25 ($M = 20.00$, SD = 1.63) years. Women of this age range are 4 times more likely to be sexually victimized compared to their counterparts (Fisher et al. 2000; Martin et al. 2022).

Key aspects of the methodology employed by Flowe et al. (2019) are presented here for convenience: Flowe et al. (2019) used a 2-beverage (tonic or alcohol) x 2-expectancy (told alcohol or told tonic) balanced placebo experimental design. Eighty women aged between 18 and 31 years ($M = 20.36$, SD = 2.41 years) were recruited from the University of Leicester and randomly assigned to a beverage and expectancy condition. Participants in the alcohol condition consumed three cups of vodka (37.5% proof) and tonic water at a 1:5 ratio. Participants in the tonic condition drank tonic water only. All beverages contained vodka-soaked limes and had vodka placed on the rims of the cups. Participant blood alcohol (BrAC) levels varied between 0.00 (i.e., the participant consumed tonic water only) and 0.12%, and mean BrAC was 0.06% (moderate intoxication). To manipulate alcohol expectancy, participants in the expect alcohol condition were told that they were consuming alcohol, whereas those in the expect tonic condition were told they were consuming tonic water.

Fifteen minutes after consuming their last beverage (which is when participants achieved peak BrAC level), they engaged in a hypothetical rape scenario. The scenario was presented using the participant choice procedure (Flowe et al. 2007), which entails asking participants to imagine themselves in a dating scenario with a male acquaintance, instructing them to think about how they would act and feel were the situation really happening to them. This allows individuals to "consent" to specific activities, which influences their perception and willingness to report rape (Flowe et al. 2007). Participants read the scenario as written text on a computer screen. They also heard the scenario being read by a female narrator through headphones.

Before beginning the scenario, participants were presented with information about the male perpetrator on a computer. This information included his physical appearance, his occupation, and his possessions, and was accompanied by a head and shoulder colour photograph of him. The hypothetical rape scenario was then presented to participants in stages. The participant was asked after each stage whether she wanted to continue to interact with the man or "call it a night" (i.e., stop interacting with the man) and end the scenario. The participant could optionally engage with him (e.g., accept a ride home from him; consent or not when he tries to kiss her, etc.). Once the participant made a choice, she was not able to return to an earlier stage of the scenario. For women who engaged all the way through the scenario, consensual sexual intercourse was depicted. If the participant chose to "call it a night" and withdrew before the scenario's end (i.e., consensual sexual intercourse), she learned that the man committed a legally definable act of rape against her (Sexual Offences Act 2003; i.e., nonconsensual sexual intercourse/rape). Ninety percent of women withdrew before the scenario ended, thus experiencing the hypothetical rape. Participants were then interviewed about the scenario with the cognitive interview a week later.

Immediately before the interview commenced, participants read an interview transcript that ostensibly belonged to another participant. It contained 4 misleading, 4 consistent, and 4 neutral pieces of information about the scenario to examine suggestibility effects. For clarity, we omitted these details in our analysis of the transcripts presented below. However, the results did not change even when we included them in the analysis. In the Appendix, we include the cognitive interview script used in the study. Note that we analyzed the data with and without the new information reported in the Change Order and Remember More sections of the transcript, and the results were consistent; hence, we did not exclude these sections in the analysis presented below. Further, both the free recall and the question phases include a mental imagery instruction.

3.2 | Recall Accuracy Coding

We followed the coding scheme implemented by Flowe et al. (2019) which was based on previous research (see Holliday 2003; Wright and Holliday 2007). A scenario template containing all details in the scenario was created. Interview transcripts were coded for accuracy using this template. A correct detail was one that was present in the scenario and recalled correctly (e.g., "red sofa" where this was depicted in the scenario), and an incorrect detail was one that was present in the scenario but recalled incorrectly (e.g., "black sofa" instead of "red sofa").

3.3 | Metamemory Processes Coding

We coded transcripts for verbal and paraverbal indicators thought to reflect underlying metamemory processes according to previous research (Lindholm et al. 2018; Gustafsson et al. 2019, 2022). Recall that these indicators were categorized into four groups: 1. Hedges; 2. Fillers; 3. Non-word fillers; and 4. Pauses (see Table 1 for operationalisation and examples). We incorporated additional examples from the linguistic literature (e.g., "just" as a hedge; Neary-Sundquist 2013) to enhance our coding scheme. Following Gustafsson et al. (2019), who found that pauses during responses were more informative of recall accuracy than response time alone, we combined response time and within-response pauses into a single "pause" category. This approach aligns with previous work by Lindholm et al. (2018) and Gustafsson et al. (2022).

While previous studies coded indicators for entire statements (Lindholm et al. 2018; Gustafsson et al. 2019, 2022), we adopted a more granular approach by coding indicators for individual details. This method allows for a more precise analysis of the relationship between metamemory processes and specific memory content. Indicators were coded for the immediately following detail within a sentence (e.g., "um he had brown hair", with "um" as the indicator and "brown" as the coded detail). In cases of multiple details following an indicator, we tagged only the first detail (e.g., "um he had dark brown hair", with "um" as the indicator, and "dark" as the first coded detail).

Coding was context-dependent, requiring coders to use judgment in classifying words as indicators (Paulo et al. 2018). For instance, "like", as in "he came up to me and was like hello", was not considered a filler as it serves to indicate reported speech. We also tallied control details, or instances where a detail was reported without preceding indicators.

Indicator data were collapsed into three categories, including no indicators, single indicators (one indicator before a detail), and multiple indicators (two or more indicators before a detail). This categorization allowed us to examine potential "dose-dependent" effects of metamemory processes on recall accuracy. We hypothesize that multiple indicators may reflect more effortful metamemory processes, potentially signaling weaker memory traces, which could be associated with lower accuracy. Table 2 presents descriptive statistics for these indicator categories across accuracy levels, interview phases, and beverage conditions.

3.4 | Phase Coding

We scored the first phase in which a given detail was reported by the participant. For instance, if the participant recalled a given piece of information in both the free recall and the question phase, the detail was scored as having been first reported in the free recall phase and was not counted again in the question phase (Prescott et al. 2011).

3.5 | Reliability Coding for Behavioral Indicators

One researcher coded 100% of the transcripts for retrieval effort indicators ($n = 22$). Fifty percent of the transcripts were randomly selected using the =*RAND()* function in Excel and coded by a second independent researcher ($n = 11$). Both researchers were blind to participants' beverage and expectancy conditions but could not be blind to the interview phase because interview phases were clear from the interview transcripts they coded. Regular coding meetings took place to discuss any queries and inconsistencies, which were resolved prior to analysis. We used Cohen's Kappa (κ) to assess agreement for the type of indicators coded before a detail. The agreement for all indicator categories was high; non-word fillers, $\kappa = 0.92$; word fillers, $\kappa = 0.91$; hedges, $\kappa = 0.81$; and pause hedges, $\kappa = 0.91$. After collapsing over indicators, reliability for single occurrences of indicators was also high; $\kappa = 0.88$; and multiple indicators was $\kappa = 0.87$. Reliability for no occurrence of indicators was $\kappa = 0.92$.

4 | Results

4.1 | Preliminary Analysis

We conducted two-proportion z-tests to investigate whether the proportion of details reported with indicator(s) (collapsed over single and multiple) significantly differed between the free recall and question phases. For both beverage conditions, the proportion of details reported with indicator(s) was significantly lower in the free recall phase compared to the question phase (0.37 versus 0.55, respectively), $z = -5.14$, $p < 0.001$. This was also the case for the tonic condition (0.35 versus 0.58, respectively), $z = -4.78$, $p < 0.001$, and the alcohol condition (0.39 versus 0.51, respectively), $z = -2.36$, $p = 0.018$.

4.2 | Data Analysis Overview

We ran a series of multilevel logistic regressions using the *glmer*[1] function from the *lme4* package (Bates et al. 2015) in R Studio (R Core Team 2018). We expected random variation within participants due to individual differences in memory performance. Data were therefore organized as multi-level data, with individual responses nested within participants (Wright and London 2009).

The outcome variable was detail accuracy (coded as 0 = incorrect, 1 = correct). The predictor variables were beverage condition (0 = tonic, 1 = alcohol), alcohol expectancy (0 = tonic, 1 = alcohol), interview phase (0 = free recall, 1 = question phase), and occurrence of indicators (no occurrence, single occurrence, multiple occurrences). Reference cell coding was used to analyze indicator type, with the no indicator category serving as the reference category.

TABLE 2 | Frequency of correct and incorrect details for each indicator type across interview phases for each beverage condition.

			Frequency				
Condition	Interview phase	Indicator	Total details recalled	Correct details	Incorrect details	Proportion	Proportion correct
Tonic ($n=10$)	Free recall	No indicators	214	191	23	0.65	0.89
		Single indicators	76	64	12	0.16	0.84
		Multiple indicators	38	35	3	0.08	0.92
	Question phase	No indicators	61	55	6	0.42	0.90
		Single indicators	44	38	6	0.30	0.86
		Multiple indicators	41	30	11	0.16	0.73
Alcohol ($n=12$)	Free recall	No indicators	167	152	15	0.61	0.91
		Single indicators	60	52	8	0.22	0.87
		Multiple indicators	49	43	6	0.18	0.88
	Question phase	No indicators	84	61	23	0.49	0.73
		Single indicators	53	35	18	0.31	0.66
		Multiple indicators	36	21	15	0.21	0.58
Total ($N=22$)	Free recall	No indicators	381	343	38	0.63	0.90
		Single indicators	136	116	20	0.23	0.85
		Multiple indicators	87	78	9	0.14	0.90
	Question phase	No indicators	145	116	29	0.45	0.80
		Single indicators	97	73	24	0.30	0.75
		Multiple indicators	77	51	26	0.24	0.66

We used the Wald Z test to assess the significance of individual predictors within each model. We used the likelihood ratio test to compare and assess model fit, as this is the most liberal test and is used frequently in eyewitness research (Mansour et al. 2017; Wright and London 2009). Alpha was set to 0.05 for all analyzes. Our hypotheses were directional and therefore were tested using one-tailed tests. All other statistical comparisons for which we had not made any predictions were two-tailed.

4.3 | Multi-Level Modeling Analysis

Table 3 shows the model parameter estimates and fit indices for all models. We first created a baseline, intercept-only model predicting accuracy (Model 1). We assessed the intraclass correlation (ICC) for our random fixed intercept model (Accuracy ~1 + (1|Participant)) and found that 5.1% of the variation in recall accuracy performance was due to differences between participants.

TABLE 3 | Parameter estimates for predictors in models 1–3 (923 observations).

Predictor	Model 1 (intercept model)	Model 2 (fixed effects model)	Model 3 (interactions model)
Fixed effects			
Intercept	**1.70 (0.13)***	**2.45 (0.25)***	**2.27 (0.29)***
Beverage		−0.37 (0.26)	0.13 (0.35)
Expectancy		0.09 (0.26)	−0.23 (0.36)
Phase		**−0.91 (0.19)***	−0.15 (0.41)
Single		**−0.38 (0.22)***	−0.61 (0.40)
Multiple		**−0.50 (0.24)***	−0.27 (0.52)
Beverage*Phase			**−1.13 (0.41)****
Beverage*Single			−0.03 (0.46)
Beverage*Multiple			−0.01 (0.52)
Expectancy*Phase			0.15 (0.41)
Expectancy*Single			0.54 (0.46)
Expectancy*Multiple			0.55 (0.52)
Phase*Single			0.03 (0.47)
Phase*Multiple			**−0.92 (0.53)***
Random parameters			
Level 2 intercept variance (participant)	0.18 (0.42)	0.15 (0.39)	0.12 (0.35)
Model fit			
Model df	2	7	15
Test change in df		5	8
AIC	803.0	778.2	780.3
BIC	812.7	812.0	852.7
−2 log likelihood	−499.5	−382.1	−375.1
Marginal R^2/conditional R^2	0.00/0.05	0.08/0.12	0.10/0.13

Note: Significant predictors are boldfaced. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Abbreviations: AIC, akaike information criterion; *B*, unstandardised logistic coefficients; BIC, Bayesian; df, degrees of freedom; Standard errors for fixed effects and standard deviations for random effects are given in parentheses.

We added beverage, expectancy, phase, single occurrences of indicators, and multiple occurrences of indicators to Model 1 to create a fixed effect model (Model 2[2]). Here, we expected phase to be a significant predictor of accuracy (Hypothesis 1). In keeping with this prediction, phase was significant, $B = -0.91$, Wald $Z = -4.75$, $p < 0.001$, one-tailed, $OR = 0.40$, (95% CI [0.28–0.59]), indicating that accuracy was significantly lower during the question compared to the free recall phase (see Figure 2). We also expected that indicators would be significant predictors of accuracy (Hypothesis 3). In keeping with this prediction, both single occurrences of indicators, $B = -0.38$, Wald $Z = -1.71$, $p = 0.044$, one-tailed, OR = 0.69, (95% CI [0.44–1.06]), and multiple occurrences of indicators, $B = -0.50$, Wald $Z = -2.05$, $p = 0.020$, one-tailed, OR = 0.61, (95% CI [0.38–0.98]) were significant. Specifically, accuracy was significantly lower with the occurrence of single and multiple indicators compared to no indicators (see Figure 3a,b). To test whether this difference between single and multiple occurrences was significant, we ran a linear hypothesis test. The results indicated no significant difference in the strength of either predictor, $\chi^2(1) = 0.22$, $p = 0.636$. Beverage, Wald $Z = -1.44$, $p = 0.151$, two-tailed, $OR = 0.69$, (95% CI [0.42–1.14]); and expectancy, Wald $Z = 0.35$, $p = 0.725$, two-tailed, $OR = 1.09$, (95% CI [0.66–1.81]) were not significant predictors of accuracy.

A likelihood ratio test indicated that the model with the fixed effects (Model 2) fit the data better compared to the intercept model (Model 1), $\chi^2(5) = 34.86$, $p < 0.001$, indicating that adding the fixed effects to Model 2 better explained the variance in accuracy compared to Model 1 (8.3% marginal R^2, 12.5% conditional R^2; Nakagawa and Schielzeth 2013).

We then added fixed interaction terms between all predictors of interest to Model 2 to create an interaction model (Model 3[2]). Here, we expected an interaction between phase and beverage, with alcohol-intoxicated participants having significantly lower accuracy during the question phase (Hypothesis 2). In line with this prediction, the interaction between phase and beverage condition was significant, $B = -1.13$, Wald $Z = -2.77$, $p = 0.003$, one-tailed, $OR = 0.32$, (95% CI [0.14–0.72]) indicating that accuracy differed significantly between phases between conditions. Figure 4 illustrates that while accuracy was similar across beverage conditions during the free recall phase, it was significantly lower for those in the alcohol condition compared to the tonic condition in the question phase. We also expected that retrieval effort indicators would predict recall accuracy, particularly in the question phase, where memories that remain to be volunteered are relatively weaker (Hypothesis 4). Our results were partially in line with this hypothesis: There was a significant interaction between phase and multiple occurrences of indicators, $B = -0.92$, Wald $Z = -1.74$, $p = 0.041$, one-tailed, $OR = 0.40$, (95% CI [0.14–1.12]). Figure 5b illustrates that accuracy was significantly lower with the occurrence of multiple indicators in the question phase compared to the free recall phase. However, there was no interaction between single occurrences of indicators and phase, Wald $Z = 0.07$, $p = 0.474$, one-tailed, $OR = 1.03$, (95% CI [0.41–2.57]). There was no interaction between expectancy and phase, Wald $Z = 0.36$, $p = 0.719$, two-tailed, $OR = 1.16$, (95% CI [0.52–2.60]), expectancy and single occurrences, Wald $Z = 1.17$, $p = 0.240$, two-tailed, $OR = 1.72$, (95% CI [0.69–4.28]); expectancy and multiple occurrences, Wald $Z = 1.07$, $p = 0.283$, two-tailed, $OR = 1.75$, (95% CI [0.63–4.86]); beverage and single occurrences, Wald $Z = -0.06$, $p = 0.953$, two-tailed, $OR = 0.97$, (95% CI [0.39–2.41]); or beverage and multiple occurrences, Wald $Z = -0.01$, $p = 0.992$, two-tailed, $OR = 0.99$, (95% CI [0.36–2.77]). We made no predictions for these interactions.

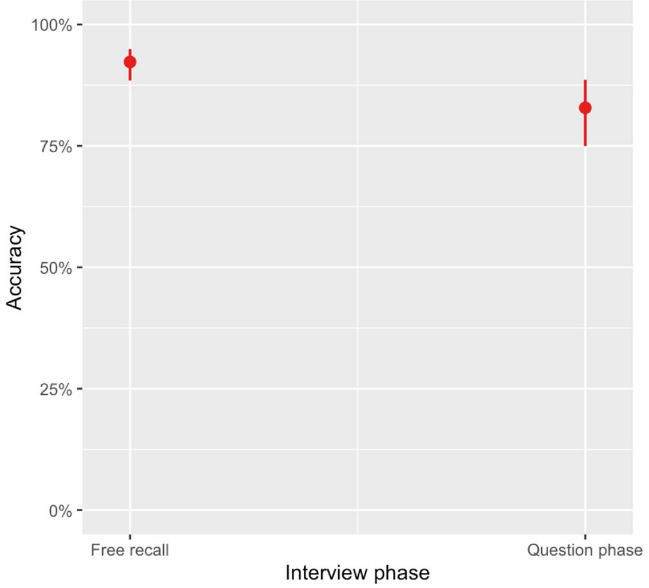

FIGURE 2 | Mean accuracy (±1 SEM) as a function of interview phase.

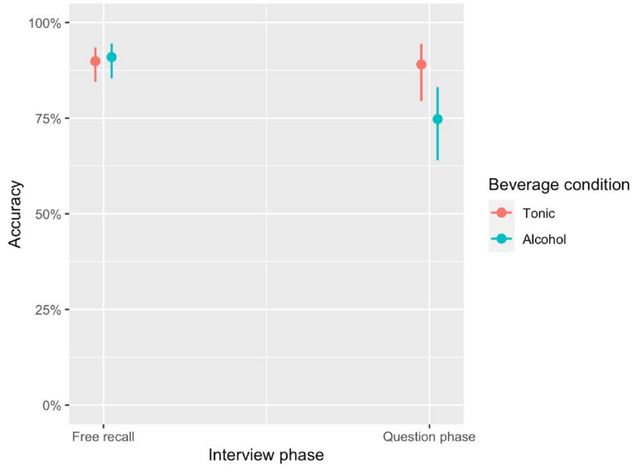

FIGURE 4 | Mean accuracy (±1 SEM) as a function of beverage condition and interview phase.

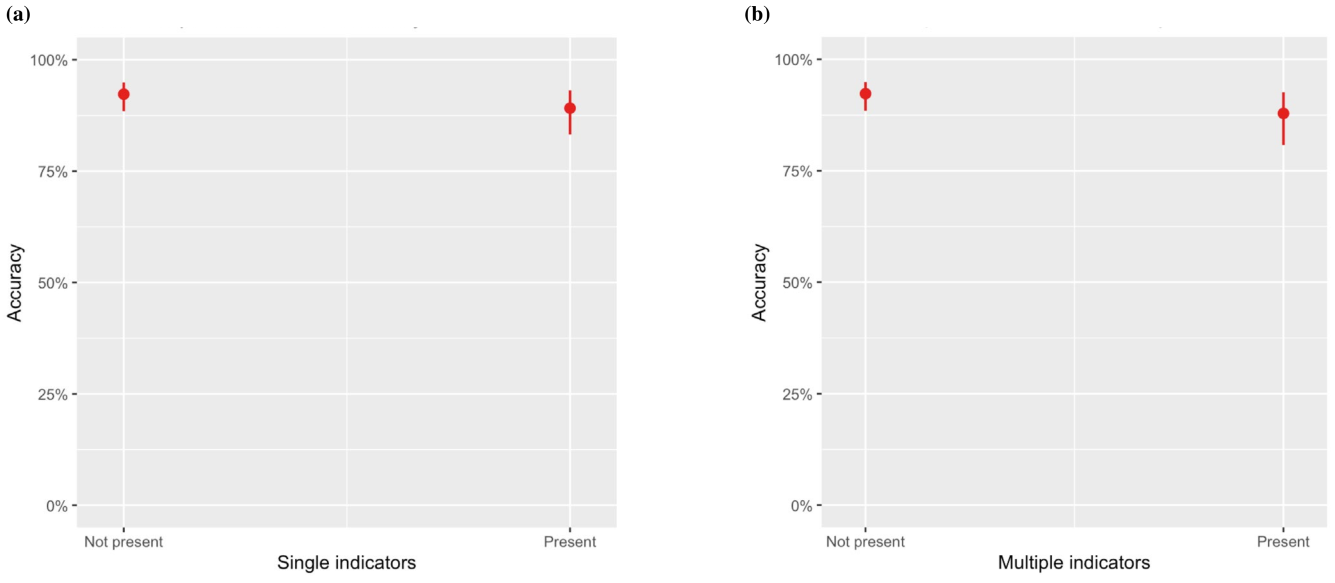

FIGURE 3 | (a) Mean accuracy (±1 SEM) as a function of single indicator occurrence. (b) Mean accuracy (±1 SEM) as a function of multiple indicator occurrences.

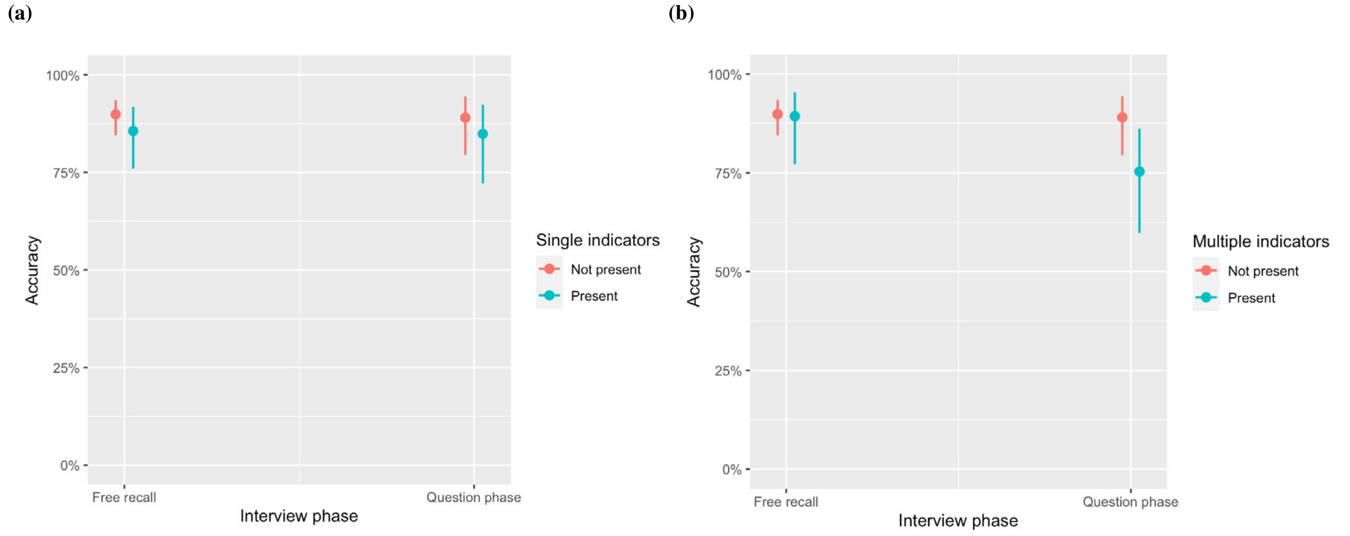

FIGURE 5 | (a) Mean accuracy (±1 SEM) as a function of single indicator occurrence and interview phase. (b) Mean accuracy (±1 SEM) as a function of multiple indicator occurrence and interview phase.

A likelihood ratio test indicated that the model with the fixed interaction terms (Model 3) did not fit the data better compared to the fixed effects model (Model 2), $\chi^2(8) = 13.87$, $p = 0.085$, suggesting that adding the fixed interaction terms in Model 3 did not better explain the variability in accuracy compared to Model 2 (9.6% marginal R^2, 12.9% conditional R^2; Nakagawa and Schielzeth 2013).

4.4 | Post Hoc Power Analysis

While our sample size of 22 participants may appear small, it is important to note that our analysis was conducted at the level of individual details ($N = 923$) rather than participants. This approach, employing multilevel modeling, allows for a more nuanced examination of the data and can provide robust results even with fewer participants. To rigorously assess the adequacy of our statistical power, we conducted post hoc power analyzes using the simR package (Green and MacLeod 2016) in R. We ran 1000 simulations based on our observed data for each analysis. Model comparisons were evaluated using likelihood ratio tests.

For the comparison between Model 1 (intercept-only) and Model 2 (fixed effects), our analysis revealed 100% power (95% CI [99.63, 100.0]). The comparison between Model 2 and Model 3 (interaction effects) yielded 77.50% power (95% CI [74.78, 80.05]). These results indicate that our dataset was sufficiently powered for these model comparisons, with the latter approaching the conventional 80% threshold. We also examined power for significant interaction terms within Model 3, following the method outlined by Stevens et al. (2023). Using 1000 simulations of a Wald Z test, we found that the beverage × phase interaction had 79.20% power (95% CI [76.55, 81.17]), again approaching the conventional threshold. The phase × multiple indicator interaction, while statistically significant in our analysis, showed lower power at 43.00% (CI [39.91, 46.14]). These power analyzes suggest that our study was generally well powered to detect the main effects and primary interactions of interest.

5 | Discussion

Our investigation into the strategic regulation of memory reporting during cognitive interviews revealed complex interactions between metamemory processes, alcohol intoxication, and interview phase. Consistent with our theoretical framework (see Figure 1), we observed a significant decline in recall accuracy from the free recall to the question phase (Hypothesis 1) that was more pronounced in participants who were alcohol intoxicated during encoding (Hypothesis 2). This pattern suggests a shift in reporting criteria as participants attempted to meet perceived demands for additional information in the question phase. The presence of metamemory indicators, particularly multiple indicators, was associated with lower recall accuracy (Hypothesis 3), especially in the question phase (Hypothesis 4). These findings show how alcohol affects not only memory encoding but also the efficacy of metamemory processes in reducing recall errors during retrieval.

Taken together, the results extend our understanding of how metamemory processes operate in police interviews, particularly under conditions of moderate alcohol intoxication during encoding. The findings highlight the need for further research into the complex interplay between memory processes, intoxication, and interview dynamics. We discuss the theoretical and applied implications of our findings next.

5.1 | Metamemory and Phased Interviews

Our results support the notion that when probed for more information during the follow-up question phase, witnesses lower their report criterion to comply with the interviewer's request for more information (see Figure 1). As a result, participants volunteer more incorrect details in the question compared to the free recall phase, as they now surpass the criterion threshold for reporting. Therefore, the metamemory processes that otherwise would have enabled the witness to withhold reporting weaker memories become less effective for maintaining accuracy during the interview. Our results are in line with previous research that found high accuracy rates when participants are allowed to freely recall details following an open prompt (Kontogianni et al. 2020; Roberts and Higham 2002), highlighting the importance of allowing witnesses to control their memory reporting (Fisher 2010). Probing for further details beyond the free recall using repetitive questioning styles may also implicitly communicate to the witness that their initial response was not adequate, possibly leading witnesses to change their answers over time, thereby increasing errors.

5.2 | Alcohol and Memory Reporting

Our study also investigated metamemory processes when the witness was alcohol intoxicated at the time of the crime. As sexual assault frequently coincides with alcohol intoxication (Abbey 2002; Hagsand et al. 2022; Palmer et al. 2013a) it is important to include alcohol as a variable in research examining rape victim memory reporting. Research has found that individuals who are intoxicated compared to sober at encoding may provide less complete statements about a crime; however, the statements provided are not less accurate (Jores et al. 2019). We found that participants who were moderately alcohol intoxicated during encoding (BrAC = 0.06) were as accurate as those in the tonic water condition for the free recall phase of the interview; however, during the question phase, their accuracy was significantly lower. Individuals may be aware that alcohol intoxication at encoding may weaken their memory and may consider this when reporting their memories. Research has found that people use "theory-based" information about how they think memory works (Palmer et al. 2013b) to strategically regulate their memory reports (see Flowe et al. 2017 for an application to alcohol and lineup identification).

This study extends previous findings in the alcohol literature on metamemory. Research has found that participants who were moderately intoxicated at encoding and sober during the lineup test can successfully use confidence to reflect their accuracy for ID lineups (Flowe et al. 2017; Sauerland et al. 2019). Similarly, research shows that participants who were alcohol intoxicated compared to sober when they encoded the crime are more likely to state "I don't know" to the interviewer's questions (Crossland et al. 2016; Flowe et al. 2018), again suggesting that alcohol-intoxicated individuals are capable of effectively monitoring and controlling their memory reporting to maintain accuracy. We extend this previous work by showing that while those who were intoxicated can strategically regulate their memory reporting to maintain accuracy during free recall, continued questioning can result in less optimal memory monitoring and control, ultimately decreasing recall accuracy.

5.3 | Metamemory Indicators

Previous research has investigated which individual indicators are useful in predicting recall accuracy (Gustafsson et al. 2019, 2022; Lindholm et al. 2018). Our results add a novel contribution to this body of research, whereby we found "dose dependent" effects of retrieval effort on accuracy. We found that the occurrence of both single and multiple metamemory indicators, as opposed to no indicator, before a detail is recalled is diagnostic of detail accuracy.

We also investigated the predictive value of indicators across interview phases as a function of beverage consumed. Incorrect details were more likely to be accompanied by multiple indicators than by no indicators across both phases, but especially so for intoxicated participants. This further supports the notion that interviewees with weakened memories may be particularly likely to lower their report criterion and volunteer weaker memories. This aligns with previous research finding that intoxicated witnesses may be more prone to make memory errors in response to repeated questioning (Oorsouw and Merckelbach 2012) and repeated suggestive questioning (Hildebrand Karlén et al. 2014).

5.4 | Implications of the Findings

First, the results suggest that investigators should perhaps view information acquired during the question phase with caution, as it may reflect a shift in metamemory strategy rather than an improvement in memory access. Second, investigators should perhaps weigh the costs and benefits of asking the victim to elaborate on their free recall account. We have learned through our interactions with legal practitioners that the accuracy of all the victim's previous statements may be called into question (by the police or at trial) if subsequent investigation shows that a given piece of information provided is incorrect no matter how tangential it may be to the investigation (e.g., the victim's ability to remember the colour of the curtains in the room where the rape occurred). Critical to this point, empirical research has demonstrated that alcohol intoxication at encoding does not affect the number of correct central details recalled (e.g., perpetrator's appearance), but does affect the number of correct peripheral details (e.g., surroundings), with intoxicated participants recalling fewer correct peripheral details than their sober counterparts (Jaffe et al. 2019; see Jores et al. 2019 for a metanalysis; Schreiber Compo et al. 2011). Given these considerations, it is crucial for investigators to critically assess the relevance of follow-up questions before asking them. This approach balances the need for comprehensive information with the importance of maintaining the overall credibility of the victim's account.

Third, research is needed to find ways to ask follow-up questions without compromising the accuracy of the account, as, of course, one can imagine many circumstances where ascertaining additional information and following-up with further questions is essential to the investigation (Evans and Fisher 2011; Shepherd and Griffiths 2013). Research that might be fruitful in this regard could include studying the effectiveness of instructions that encourage interviewees to maintain their reporting criterion (i.e., not to volunteer information they had previously withheld during their initial free recall) when answering focused questions, or instructions to the interviewee to adjust the grain size (i.e., to provide "coarse-grained" information, like "jacket", rather than "fine-grained" information, like "black leather jacket") of the details that they report (Brewer et al. 2018) to help the interviewee maintain accuracy. Evans and Fisher (2011) suggest that specific reporting instructions may help the witness balance accuracy and informativeness demands for follow-up questions (cf. Kontogianni et al. 2020). Another intervention that might prove to be effective is to encourage interviewers to explicitly acknowledge during the interview that they are asking questions that are speculative or that may be difficult or impossible for the victim to answer based on their previous statements. The transfer of control instructions in the modified cognitive interview tell participants to say "I don't know" ("it is OK to say they don't know or are unsure about something"), that the interviewer does not share their knowledge ("I don't know what happened"), and to correct the interviewer if they say something wrong ("if I say something that's wrong, just tell me I'm wrong"). Explicit acknowledgement of speculation in addition to these may help to ameliorate the potential damage that incorrect answers to these types of questions could have on perceptions of the victim's credibility and the accuracy of other aspects of their account.

5.5 | Limitations

There are several limitations to consider with this type of research. First, the level of alcohol intoxication (BrAC = 0.06) employed is lower than what is typically encountered in forensic settings (Flowe and Carline 2021). While generalization to higher intoxication levels requires empirical study, our theoretical framework predicts that the observed effects on metamemory processes and memory reporting would be greater at higher blood alcohol concentrations. This prediction stems from the established dose-dependent nature of alcohol's impact on cognition and memory (e.g., Bisby et al. 2010; Weissenborn and Duka 2000). Future research examining higher intoxication levels could provide valuable insights into the scalability of the effects we observed.

Second, the interviews we analyzed were conducted in a university research context, rather than in a police environment. The external validity of this type of research depends on the extent to which psychological realism is achieved by the ability of the experimental procedures to instantiate the psychological processes that are most theoretically relevant. Third, we analyzed interview transcripts from a study that used a hypothetical rape scenario to test remembering. While this enabled us to measure memory accuracy, the scenario methodology does not fully capture the complexity and traumatic impact of real sexual assault experiences on memory. The method of scenario presentation aimed to achieve psychological realism (Mook 1983). Participants in other studies experiencing this scenario paradigm have found it distressing, and they reported mild traumatic stress symptoms (Takarangi et al. 2013, 2016). Having said this, it is still necessary to triangulate experimental findings with the experiences of survivors and legal practitioners.

Fourth, Flowe et al. (2019) did not collect information about participants' sexual assault victimization because they did not have any hypotheses about how such experiences would moderate the effects of alcohol on memory recall. While we know of no theoretical basis to suggest this variable would influence the relationship between alcohol and memory recall, future research might explore whether and how previous experiences might affect this relationship. Fifth, behavioral indicators are an indirect measure of metamemory processes the can be influenced by other factors besides memory strength. As noted by previous researchers (Budescu and Wallsten 1995; Gustafsson et al. 2022; Erev and Cohen 1990; Fontaine 2017; Lindholm et al. 2018), the use of retrieval effort indicators, and the specific types of indicators used, may differ between individuals and social contexts, making it difficult to advise what types of indicators are the best for indexing retrieval effort. The indicators we analyzed may have been affected by other psychological processes, such as nervousness (Goberman et al. 2011) and deception (Sporer and Schwandt 2006), as well as conversational dynamics (Clark and Tree 2002). Task anxiety may also interfere with metamemory processes (Veenman et al. 2000). Given that victims of sexual violence often face challenges to their memory's veracity (Antaki et al. 2015; MacLeod 2020) and experience disbelief (Greeson et al. 2016), future research must examine how anxiety and these contextual factors, such as interviewer behavior, influence metamemory measurement.

Finally, replication with a larger sample size would be valuable to confirm and extend these findings, particularly for the interaction effects we observed. Power analyzes in multilevel models are complex, and traditional benchmarks may not always apply straightforwardly (Green and MacLeod 2016). Notably, we found a significant interaction between alcohol consumption and interview phase on recall accuracy, whereas Flowe et al. (2019) observed only a non-significant trend. They used a traditional mixed-effects ANOVA, which, while robust, may be less sensitive to detecting interactions with repeated measures compared to our linear mixed-effects approach (Baayen et al. 2008). This underscores the increased sensitivity of multilevel modeling in capturing subtle but meaningful effects, even with a relatively small number of participants.

5.6 | Conclusions

Our findings suggest that extensive questioning contributes to inaccuracies in victim statements, potentially exacerbating the challenges faced by sexual assault survivors in the criminal justice system. This is particularly problematic with alcohol-intoxicated interviewees. Our findings have significant implications for cases like Chanel Miller's, where alcohol intoxication and incomplete memories are often used to discredit the accuracy of statements and testimony of sexual assault victims as well as their credibility. Further research is needed to better understand metamemory processes during interviews and to develop more effective interview approaches in such cases.

Author Contributions

Madeleine P. Ingham: investigation, methodology, visualization, writing – review and editing, writing – original draft, data curation, formal analysis. **Brittany D. Gibbs:** data curation, formal analysis, investigation, methodology. **Melissa F. Colloff:** investigation, methodology, data curation, writing – review and editing, validation. **Laura M. Stevens:** data curation, formal analysis, writing – review and editing. **Orli M. Edwards:** data curation, writing – review and editing.

Sarah R. Rockowitz: data curation, writing – review and editing. **Rumandeep K. Hayre:** formal analysis, writing – review and editing. **Mussaffa Butt:** data curation, writing – review and editing. **Chloe A. Morris:** visualization. **Heather D. Flowe:** conceptualization, investigation, funding acquisition, writing – review and editing, writing – original draft, methodology, validation, visualization, data curation, formal analysis, supervision, project administration, resources.

Ethics Statement

The Flowe et al. (2019) study protocol was approved by the University of Leicester School of Psychology Ethics Committee. The current study was ethically approved by the University of Birmingham Science, Technology, Engineering and Mathematics Ethical Review Committee.

Consent

All participants provided their written informed consent to participate in this study. Participants were informed of their right to withdraw at any time, and their data was anonymised for confidentiality.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] For all models we used the *bobyqa* optimiser and 20,000 function evaluations.

[2] Assumptions (residual uniformity, dispersion, heteroscedasticity, outliers and zero inflation and multicollinearity) were checked using the car (Fox and Weisberg 2019) and DHARMa (Hartig 2022) packages in R studio. All assumptions were met.

References

Abbey, A. 2002. "Alcohol-Related Sexual Assault: A Common Problem Among College Students." *Journal of Studies on Alcohol, Supplement* 14: 118–128.

Abbey, A., T. Zawacki, P. O. Buck, A. M. Clinton, and P. McAuslan. 2004. "Sexual Assault and Alcohol Consumption: What Do We Know About Their Relationship and What Types of Research Are Still Needed?" *Aggression and Violent Behavior* 9, no. 3: 271–303. https://doi.org/10.1016/S1359-1789(03)00011-9.

Ackerman, R., and M. Goldsmith. 2008. "Control Over Grain Size in Memory Reporting—With and Without Satisficing Knowledge." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34, no. 5: 1224–1245.

Antaki, C., E. Richardson, E. Stokoe, and S. Willott. 2015. "Police Interviews With Vulnerable People Alleging Sexual Assault: Probing Inconsistency and Questioning Conduct." *Journal of SocioLinguistics* 19, no. 3: 328–350. https://doi.org/10.1111/josl.12124.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. "Mixed-Effects Modelling With Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59, no. 4: 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67, no. 1: 1–48. https://doi.org/10.18637/jss.v067.i01.

Bisby, J. A., J. R. Leitz, C. J. Morgan, and H. V. Curran. 2010. "Decreases in Recollective Experience Following Acute Alcohol: A Dose–Response Study." *Psychopharmacology* 208, no. 1: 67–74. https://doi.org/10.1007/s00213-009-1709-y.

Brewer, N., A. N. Vagadia, L. Hope, and F. Gabbert. 2018. "Interviewing Witnesses: Eliciting Coarse-Grain Information." *Law and Human Behavior* 42, no. 5: 458–471. https://doi.org/10.1037/lhb0000294.

Budescu, D. V., and T. S. Wallsten. 1995. "Processing Linguistic Probabilities: General Principles and Empirical Evidence." In *Psychology of Learning and Motivation*, vol. 32, 275–318. Academic Press.

Clark, H. H., and J. E. F. Tree. 2002. "Using Uh and Um in Spontaneous Speaking." *Cognition* 84, no. 1: 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3.

College of Policing. 2022. "Investigative Interviewing." https://www.college.police.uk/app/investigation/investigative-interviewing.

Crossland, D., W. Kneller, and R. Wilcock. 2016. "Intoxicated Witnesses: Testing the Validity of the Alcohol Myopia Theory." *Applied Cognitive Psychology* 30, no. 2: 270–281. https://doi.org/10.1002/acp.3209.

Davis, D., and E. F. Loftus. 2015. "Remembering Disputed Sexual Encounters: A New Frontier for Witness Memory Research." *Journal of Criminal Law and Criminology* 105, no. 4: 811–852.

Erev, I., and B. L. Cohen. 1990. "Verbal Versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox." *Organizational Behavior and Human Decision Processes* 45, no. 1: 1–18.

Evans, J. R., and R. P. Fisher. 2011. "Eyewitness Memory: Balancing the Accuracy, Precision and Quantity of Information Through Metacognitive Monitoring and Control." *Applied Cognitive Psychology* 25, no. 3: 501–508. https://doi.org/10.1002/acp.1722.

Evans, J. R., and N. Schreiber Compo. 2010. "Mock Jurors' Perceptions of Identifications Made by Intoxicated Eyewitnesses." *Psychology, Crime & Law* 16, no. 3: 191–210. https://doi.org/10.1080/10683160802612890.

Evans, J. R., N. Schreiber Compo, and M. B. Russano. 2009. "Intoxicated Witnesses and Suspects: Procedures and Prevalence According to Law Enforcement." *Psychology, Public Policy, and Law* 15, no. 3: 194–221. https://doi.org/10.1037/a0016837.

Fisher, B. S., F. T. Cullen, and M. G. Turner. 2000. *The Sexual Victimization of College Women (NCJ 182369 (Open in a New Window))*. U.S. Department of Justice, Bureau of Justice Statistics and National Institute of Justice.

Fisher, R. P. 2010. "Interviewing Cooperative Witnesses." *Legal and Criminological Psychology* 15, no. 1: 25–38. https://doi.org/10.1348/135532509X441891.

Flowe, H., J. Humphries, M. Takarangi, et al. 2019. "An Experimental Examination of the Effects of Alcohol Consumption and Exposure to Misleading Post Event Information on Remembering a Hypothetical Rape Scenario." *Applied Cognitive Psychology* 33, no. 3: 393–413. https://doi.org/10.1002/acp.3531.

Flowe, H. D., and A. Carline. (eds.) 2021. "Alcohol and Remembering Rape: Setting the Scene." In *Alcohol and Remembering Rape: New Evidence for Practice*, 1–19. Palgrave-McMillan. https://doi.org/10.1007/978-3-030-67867-8.

Flowe, H. D., A. Carline, and N. Karoğlu. 2018. "Testing the Reflection Assumption: A Comparison of Eyewitness Ecology in the Laboratory and Criminal Cases." *International Journal of Evidence and Proof* 22, no. 3: 239–261. https://doi.org/10.1177/1365712718782996.

Flowe, H. D., M. F. Colloff, K. Zelek, J. E. Humphries, and M. K. T. Takarangi. 2017. "The Effects of Alcohol Intoxication on Accuracy and the Confidence-Accuracy Relationship in Photographic Simultaneous Lineups." *Applied Cognitive Psychology* 31, no. 4: 379–391. https://doi.org/10.1002/acp.3332.

Flowe, H. D., E. B. Ebbesen, and A. Putcha-Bhagavatula. 2007. "Rape Shield Laws and Sexual Behavior Evidence: Effects of Consent Level and Women's Sexual History on Rape Allegations." *Law and Human Behavior* 31, no. 2: 159–175. https://doi.org/10.1007/s10979-006-9050-z.

Fontaine, E. L. 2017. "Narrative Recall in an Investigative Interview: Insight Into Witness Metacognition." Doctoral diss.

Fox, J., and S. Weisberg. 2019. *Using Car Functions in Other Functions*. CRAN R.

Goberman, A. M., S. Hughes, and T. Haydock. 2011. "Acoustic Characteristics of Public Speaking: Anxiety and Practice Effects." *Speech Communication* 53, no. 6: 867–876. https://doi.org/10.1016/j.specom.2011.02.005.

Goldsmith, M., A. Koriat, and A. Pansky. 2005. "Strategic Regulation of Grain Size in Memory Reporting Over Time." *Journal of Memory and Language* 52, no. 4: 505–525. https://doi.org/10.1016/j.jml.2005.01.010.

Goldsmith, M., A. Koriat, and A. Weinberg-Eliezer. 2002. "Strategic Regulation of Grain Size Memory Reporting." *Journal of Experimental Psychology: General* 131, no. 1: 73–95. https://doi.org/10.1037/0096-3445.131.1.73.

Green, P., and C. J. MacLeod. 2016. "SIMR: An R Package for Power Analysis of Generalized Linear Mixed Models by Simulation." *Methods in Ecology and Evolution* 7, no. 4: 493–498. https://doi.org/10.1111/2041-210X.12504.

Greeson, M. R., R. Campbell, and G. Fehler-Cabral. 2016. ""Nobody Deserves This": Adolescent Sexual Assault Victims'Perceptions of Disbelief and Victim Blame From Police." *Journal of Community Psychology* 44, no. 1: 90–110. https://doi.org/10.1002/jcop.21744.

Gustafsson, P., L. Torun, and F. Jönsson. 2019. "Predicting Accuracy in Eyewitness Testimonies With Memory Retrieval Effort and Confidence." *Frontiers in Psychology* 10, no. 703: 1–10. https://doi.org/10.3389/fpsyg.2019.00703.

Gustafsson, P. U., T. Lindholm, and F. U. Jönsson. 2022. "Eyewitness Accuracy and Retrieval Effort: Effects of Time and Repetition." *PLoS One* 17, no. 9: e0273455. https://doi.org/10.1371/journal.pone.0307692.

Hagsand, A. V., D. Pettersson, J. R. Evans, and N. Schreiber-Compo. 2022. "Police Survey: Procedures and Prevalence of Intoxicated Witnesses and Victims in Sweden." *European Journal of Psychology Applied to Legal Context* 14, no. 1: 21–31. https://doi.org/10.5093/ejpalc2022a3.

Hartig, F. 2022. "DHARMa: Residual Diagnostics for Hierarchical (multi-level/mixed) Regression Models." R Package Version 0.4. 6.

Hildebrand Karlén, M., E. Roos af Hjelmsäter, C. Fahlke, P. A. Granhag, and A. Söderpalm Gordh. 2014. "Alcohol Intoxicated Eyewitnesses' Memory of Intimate Partner Violence." *Psychology, Crime & Law* 21, no. 2: 156–171. https://doi.org/10.1080/1068316X.2014.951644.

Holliday, R. E. 2003. "The Effect of a Prior Cognitive Interview on Children's Acceptance of Misinformation." *Applied Cognitive Psychology* 17, no. 4: 443–457. https://doi.org/10.1002/acp.879.

Jaffe, A. E., C. M. Harris, and D. DiLillo. 2019. "Observing Alcohol Myopia in the Context of a Trauma Film Paradigm: Differential Recall of Central and Peripheral Details." *Alcoholism: Clinical and Experimental Research* 43, no. 10: 2203–2211.

Jores, T., M. F. Colloff, L. Kloft, H. Smailes, and H. D. Flowe. 2019. "A Meta-Analysis of the Effects of Acute Alcohol Intoxication on Witness Recall." *Applied Cognitive Psychology* 33, no. 3: 334–343. https://doi.org/10.1002/acp.3533.

Kontogianni, F., L. Hope, P. J. Taylor, A. Vrij, and F. Gabbert. 2020. ""Tell Me More About This…": An Examination of the Efficacy of Follow-Up Open Questions Following an Initial Account." *Applied Cognitive Psychology* 34, no. 5: 972–983. https://doi.org/10.1002/acp.3675.

Koriat, A., and M. Goldsmith. 1996. "Monitoring and Control Processes in the Strategic Regulation of Memory Accuracy." *Psychological Review* 103, no. 3: 490–517. https://doi.org/10.1037/0033-295X.103.3.490.

Lindholm, T., F. U. Jönsson, and M. T. Liuzza. 2018. "Retrieval Effort Cues Predict Eyewitness Accuracy." *Journal of Experimental Psychology: Applied* 24, no. 4: 534–542. https://doi.org/10.1037/xap0000175.

MacLeod, N. 2020. "Tell Me in Your Own Words…": Reconciling Institutional Salience and Witness-Compatible Language in Police Interviews With Women Reporting Rape.

Mansour, J. K., J. L. Beaudry, and R. C. L. Lindsay. 2017. "Are Multiple-Trial Experiments Appropriate for Eyewitness Identification Studies? Accuracy, Choosing, and Confidence Across Trials." *Behavior Research Methods* 49, no. 6: 2235–2254. https://doi.org/10.3758/s13428-017-0855-0.

Martin, S. L., B. S. Fisher, M. C. Stoner, C. F. Rizo, and M. L. Wojcik. 2022. "Sexual Assault of College Students: Victimization and Perpetration Prevalence Involving Cisgender Men, Cisgender Women and Gender Minorities." *Journal of American College Health* 70, no. 2: 404–410.

Memon, A., C. Meissner, and J. Fraser. 2010. "The Cognitive Interview: A Meta-Analytic Review and Study Space Analysis of the Past 25 Years." *Psychology, Public Policy, and Law* 16: 340–372. https://doi.org/10.1037/a0020518.

Mickes, L., T. M. Seale-Carlisle, and J. T. Wixted. 2013. "Rethinking Familiarity: Remember/Know Judgments in Free Recall." *Journal of Memory and Language* 68, no. 4: 333–349. https://doi.org/10.1016/j.jml.2013.01.001.

Miller, C. 2019. *Know My Name. A Memoir.* Penguin.

Ministry of Justice. 2022. *Achieving Best Evidence in Criminal Proceedings Guidance on Interviewing Victims and Witnesses, and Guidance on Using Special Measures*. Home Office. https://www.cps.gov.uk/sites/default/files/documents/legal_guidance/best_evidence_in_criminal_proceedings.pdf.

Mohler-Kuo, M., G. W. Dowdall, M. P. Koss, and H. Wechsler. 2004. "Correlates of Rape While Intoxicated in a National Sample of College Women." *Journal of Studies on Alcohol* 65, no. 1: 37–45. https://doi.org/10.15288/jsa.2004.65.37.

Mook, D. 1983. "In Defense of External Invalidity." *American Psychologist* 38: 379–387. https://doi.org/10.1037/0003-066x.38.4.379.

Nakagawa, S., and H. Schielzeth. 2013. "A General and Simple Method for Obtaining R2 From Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4, no. 2: 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Neary-Sundquist, C. 2013. "The Use of Hedges in the Speech of ESL Learners." *Estudios de Lingüística Inglesa Aplicada* 2013, no. 13: 149–174.

Nelson, T. O., and L. Narens. 1994. "Why Investigate Metacognition?" In *Metacognition: Knowing About Knowing*, edited by J. Metcalfe and A. P. Shimamura, 1–25. The MIT Press.

Van Oorsouw, K., and H. Merckelbach. 2012. "The Effects of Alcohol on Crime-Related Memories: A Field Study." *Applied Cognitive Psychology* 26, no. 1: 82–90. https://doi.org/10.1002/acp.1799.

Palmer, F. T., H. D. Flowe, M. K. T. Takarangi, and J. E. Humphries. 2013a. "Intoxicated Witnesses and Suspects: An Archival Analysis of Their Involvement in Criminal Case Processing." *Law and Human Behavior* 37, no. 1: 54–59. https://doi.org/10.1037/lhb0000010.

Palmer, M. A., N. Brewer, N. Weber, and A. Nagesh. 2013b. "The Confidence-Accuracy Relationship for Eyewitness Identification Decisions: Effects of Exposure Duration, Retention Interval, and Divided Attention." *Journal of Experimental Psychology: Applied* 19, no. 1: 55–71. https://doi.org/10.1037/a0031602.

Paulo, R. M., P. B. Albuquerque, and R. Bull. 2018. "The Enhanced Cognitive Interview: Expressions of Uncertainty, Motivation and Its Relation With Report Accuracy." In *Investigating, the Truth*, 76–92. Routledge.

Prescott, K., R. Milne, and J. Clarke. 2011. "How Effective Is the Enhanced Cognitive Interview When Aiding Recall Retrieval of Older Adults Including Memory for Conversation?" *Journal of Investigative Psychology and Offender Profiling* 8, no. 3: 257–270. https://doi.org/10.1002/jip.142.

R Core Team. 2018. "R: A Language and Environment for Statistical Computing." https://www.R-project.org/.

Roberts, W. T., and P. A. Higham. 2002. "Selecting Accurate Statements From the Cognitive Interview Using Confidence Ratings." *Journal of Experimental Psychology: Applied* 8, no. 1: 33–43. https://doi.org/10.1037/1076-898X.8.1.33.

Sauerland, M., N. J. Broers, and K. I. M. Van Oorsouw. 2019. "Two Field Studies on the Effects of Alcohol on Eyewitness Identification, Confidence, and Decision Times." *Applied Cognitive Psychology* 33, no. 3: 370–385. https://doi.org/10.1002/acp.3493.

Schreiber Compo, N., J. R. Evans, R. N. Carol, et al. 2011. "Alcohol Intoxication and Memory for Events: A Snapshot of Alcohol Myopia in a Real-World Drinking Scenario." *Memory* 19, no. 2: 202–210.

Sexual Offences Act. (2003). legislation.gov.uk. https://www.legislation.gov.uk/ukpga/2003/42/contents.

Shepherd, E., and A. Griffiths. 2013. *Investigative Interviewing: The Conversation Management Approach*. 2nd ed. Oxford University Press.

Söderlund, H., E. S. Parker, B. L. Schwartz, et al. 2005. "Memory Encoding and Retrieval on the Ascending and Descending Limbs of the Blood Alcohol Concentration Curve." *Psychopharmacology* 182: 305–317. https://doi.org/10.1007/s00213-005-0096-2.

Sporer, S. L., and B. Schwandt. 2006. "Paraverbal Indicators of Deception: A Meta-Analytic Synthesis." *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 20, no. 4: 421–446. https://doi.org/10.1002/acp.1190.

Stern, V. 2010. *The Stern Review: A Report by Baroness Vivien Stern CBE of an Independent Review Into How Rape Complaints Are Handled by Public Authorities in England and Wales*. Home Office.

Stevens, L. M., L. A. Monds, B. Riordan, R. K. Hayre, and H. D. Flowe. 2023. "Acute Alcohol Intoxication and Alcohol Expectancy Effects on Women's Memory for Consensual and Non-Consensual Sexual Activity." *Frontiers in Psychology* 13: 1008563. https://doi.org/10.3389/fpsyg.2022.1008563.

Stokoe, E. 2013. "The (In)authenticity of Simulated Talk: Comparing Role-Played and Actual Interaction and the Implications for Communication Training." *Research on Language &Amp; Social Interaction* 46, no. 2: 165–185. https://doi.org/10.1080/08351813.2013.780341.

Takarangi, M. K. T., H. D. Flowe, and J. E. Humphries. 2013. *The Effects of Alcohol on the Development of Intrusive Memories for Sexual Assault*. Society for Applied Research in Memory and Cognition.

Takarangi, M. K. T., R. A. Smith, D. Strange, and H. D. Flowe. 2016. "Metacognitive and Metamemory Beliefs in the Development and Maintenance of Posttraumatic Stress Disorder." *Clinical Psychological Science* 5, no. 1: 131–140. https://doi.org/10.1177/2167702616649348.

Veenman, M. V. J., L. Kerseboom, and C. Imthorn. 2000. "Test Anxiety and Metacognitive Skillfulness: Availability Versus Production Deficiencies." *Anxiety, Stress & Coping: An International Journal* 13, no. 4: 391–412. https://doi.org/10.1080/10615800008248343.

Weissenborn, R., and T. Duka. 2000. "State-Dependent Effects of Alcohol on Explicit Memory: The Role of Semantic Associations." *Psychopharmacology* 149, no. 1: 98–106. https://doi.org/10.1007/s002139900349.

Westera, N. J., and M. B. Powell. 2017. "Prosecutors' Perceptions of How to Improve the Quality of Evidence in Domestic Violence Cases." *Policing and Society* 27, no. 2: 157–172. https://doi.org/10.1080/10439463.2015.1039002.

Westera, N. J., M. B. Powell, and B. Milne. 2017. "Lost in the Detail: Prosecutors' Perceptions of the Utility of Video Recorded Police Interviews as Rape Complainant Evidence." *Australian and New Zealand Journal of Criminology* 50, no. 2: 252–268. https://doi.org/10.1177/0004865815620705.

Wright, A., and R. Holliday. 2007. "Interviewing Cognitively Impaired Older Adults: How Useful Is a Cognitive Interview?" *Memory* 15, no. 1: 17–33. https://doi.org/10.1080/09658210601047351.

Wright, D., and K. London. 2009. "Multilevel Modelling: Beyond the Basic Applications." *British Journal of Mathematical and Statistical Psychology* 62, no. 2: 439–456. https://doi.org/10.1348/000711008X327632.

Appendix A

Modified Cognitive Interview Script

1. Introductory Phase.

1a. Rapport building.

1b. Explain the aims/rules of the interview and transfer control.

Transfer of control: "I'd like you to tell me what happened in the scenario you read and heard last week. Don't make anything up or guess. It's OK to say you don't know or you are unsure about something. I don't know what happened, so if I say something that's wrong, just tell me I'm wrong. And if you don't understand something I say, tell me."

Part of this session will involve you closing your eyes. If that makes you feel uncomfortable, that's fine, I'll just need you to look down and focus on the floor instead. If it's OK with you, I'll record us talking and write down some things, just to help me remember what you say for later on. Do you have any questions?

2. Free recall phase.

2a. Context reinstatement.

"OK, so first of all, please close your eyes and picture in your mind the dating scenario you experienced last week. It might help to recall where you were in the scenario, what you visualized, what you were thinking and how you were feeling at the time." [Pause].

Visualize what happened in your mind and think about the following things:

•Where you were •What you were doing •Who you were with •How you were feeling.

•What was happening •Who was involved •What you could see and hear in your mind.

2b. Report all.

"Now I'd like you to tell me everything you can remember about the event and the people involved… even things that you think may not be important. Please give me as many details as possible, without leaving anything out, and without guessing about the information. We are only interested in your own memories of the event." [Pause; wait for response].

2c. Change order.

"OK so now I'd like to try something different that can help me to remember other information. Please can you tell me about the very last thing that you remember in the scenario" [Pause; wait for response].

"OK thank you. Now tell me what happened just before that?" [Pause; wait for response].

Continue asking this until the interviewee reaches the beginning of the scenario.

Only note down any additional information and slot it into the free recall order.

3. Remember more.

"That's great. Can you remember anything more about the scenario?"

4. Questioning phase with mental imagery.

"I'm going to ask you a few questions about what you have told me about the scenario"

Follow-up on the man/scene/car in the order that they mentioned them. If they did not mention them (e.g., *car), ask "Were there any vehicles involved?" Use their terminology* (e.g., *say "guy" if they said guy rather than man*).

E.g., "Please close your eyes again. You mentioned a man earlier. Try and picture that man in your head. Can you tell me anything more about the man?"

[Pause; wait for response].

"Without guessing, can you remember his..."

Only ask for the following details if they have not mentioned them already. Ask line by line.

• Apparent Age • Height.

• Ethnic origin • Weight/Build • Features e.g. Eyes/Ears/Mouth/Nose/ etc.

• Hair Color • Facial Hair • Complexion.

• Clothing/Shoes • Accent • Glasses.

• "Jewelry • Accessories • Scars/Marks/Tattoos".

"Can you remember anything else about the man?"

The scenarios usually involve two scenes (e.g., *the bar and her house later on). Ask about each in turn.*

"You mentioned a [bar] earlier. Try and picture the [bar] scene in your head. Can you tell me anything more about the [bar] scene?" [Pause; wait for response].

"Again, without guessing, please can you provide a description of the [bar] scene as you remember it? Please include details of where you were, where other people were, the movement of yourself and other people you saw, and also details of any features of the scene."

Depending on their response...

"At the [bar] scene, were other people present who saw what happened?"

"Can you provide a description?"

"Can you remember anything else about the [bar] scene?"

Now follow up on the second scene.

"You mentioned that you went back to [your home]. Try and picture [your home] scene in your head. Can you tell me anything more about [your home] scene?"

[Pause; wait for response].

"Again, without guessing, please can you provide a description of [your home] scene as you remember it? Please include details of where you were, where other people were, the movement of yourself and other people you saw, and also details of any features of the scene."

Depending on their response...

"At [your home] scene, were other people present who saw what happened?"

"Can you provide a description?"

"Can you remember anything else about [your home] scene?"

"You mentioned a car... Try and picture the car in your head. Can you tell me anything more about the car?"

[Pause; wait for response].

"Again, without guessing, please can you provide as much detail as you can about the car. For instance..."

• Size • Shape • Colour.

• Make/model • Number of Doors • Registration Number.

• You mentioned that he drove you back. Can you remember details such as the driving style and speed at which you were traveling?

"Can you remember anything else about the car?"

"OK, I'm just going to ask you a few more questions".

How well in your mind did you see the incident?

How long was the entire scenario? *If they ask about the split scenes, ask for each in turn and then altogether.*

What were the weather conditions like at the time?

What time of day did the event occur?

Did you view the incident in daylight or artificial light? (Describe if possible).

Are there any particular reasons for remembering the event or the man portrayed in the scenario?

Was anyone involved that you know or whom you have seen before? (If so, where and when?).

5. Closure "Do you have any questions? Thank you for your help"